Atmospheric
Measurement
Techniques
Discussions

Open Access

Applying receptor models Unmix and PMF on real data set of elements in PM for sources evaluation of the sea coastal side region (Southeast Adriatic Sea)

D. Đorđević[1], S. Petrović[2], D. Relić[3], and A. Mihajlidi-Zelić[3]

[1]ICTM – Centre of Chemistry, University of Belgrade, Studentski trg 12–16,
11000 Belgrade, Serbia
[2]ICTM – Centre of Catalysis, University of Belgrade, Studentski trg 12–16,
11000 Belgrade, Serbia
[3]Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Serbia

Correspondence to: D. Đorđević (dragadj@chem.bg.ac.rs)

AMTD

6, 4941–4969, 2013

Applying receptor models Unmix and PMF on real data set of elements in PM

D. Đorđević et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|
◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

D. Đorđević et al.

Abstract

Two receptor modeling techniques were applied to a common data set of daily measurements of 11 elements in particulate matter (PM). Samples of PM were collected in the 5 yr period at an urban site located at the sea coast. In the vicinity of the sampling site traffic is a permanent but not significant anthropogenic source. In this study we used both the Unmix and PMF receptor models for evaluation of the sources contribution. Unmix found thirteen solutions for several combinations of species, but four solutions satisfy the criteria Min $R^2 > 0.8$ and Min $S/N > 2$. Unmix identified three and four sources in satisfactory solutions. The PMF model has given 3 possible solutions and by further analysis the best solution of four sources was selected. F peak refinement enabled finding a more realistic solution that includes re-suspension and traffic as dominant source contributions. The results given in this study are in accordance with the results of Enrichment Factors analysis presented in our previous work.

1 Introduction

Several multivariate receptor models like principal component analysis (PCA) and factor analysis (FA) have been useful in providing a relatively easy and accurate means for classifying sources (Lee et al., 2003). However, because PCA and FA analysis depend only on the covariance matrix, general requirements for the source contributions are not satisfied, e.g. negative values for factor loading and factor score often occur.

The Chemical Mass Balance (CMB) method is widely used to estimate contributions of emission sources to receptors. Weighted least-squares solutions are solved to an equation, which describes a mass balance between ambient concentration of each chemical species and a linear sum of products of source profiles and source contributions (Lee et al., 2009).

A state-of-the-art multivariate receptor models are applied in the diverse fields of environmentrics, chemometrics, geology and remote sensing. Multivariate receptor

modeling is a term applied in the field of air quality for the solution of the general linear mixture problem. For conservative chemical species, i.e. those that do not undergo reactions in the atmosphere, the principle of mass balance is applied. The mass balance for species i can be written as:

$$C_{ij} = \sum_{k=1}^{N} a_{ik} S_{kj} \quad i = 1, \ldots, m, \quad j = 1, \ldots, n. \tag{1}$$

In this equation, C_{ij} is the observed concentration of species i in sample j, S_{kj} is the total amount of particulate mass from source k in sample j and a_{ik} is the composition fraction of species i from the source k. In air quality studies, the units of C_{ij} are usually micrograms per cubic meter. Thus, since a_{ik} is a dimensionless mass fraction, the units of S_k are also micrograms per cubic meter. Equation (1) is the physical basis of all receptor models. C_{ij} is subject to random error and a_{ik} to random variations (Henry, 2002).

Unmix seeks to solve the general mixture problem where the data is assumed to be a linear combination of an unknown number of sources of unknown composition, which contribute an unknown amount to each sample (Norris et al., 2007). Unmix assumes that the data and the compositions of the sources are all strictly positive (because of the effects of errors, small values less than zero are allowed in order to reduce the bias in the results). Unmix further assumes that for each source there are some samples that contain little or no contribution from that source. For a given selection of species, Unmix estimates the number of sources, the source composition, and source contributions to each sample. The usual analytical approach to fitting the model in Eq. (1) is to find the values of a_{ik} and S_{kj} that minimize the weighted mean square error F (Henry, 2002) of the model:

$$F = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(w_{ij} C_{ij} - \sum_{k=1}^{N} a_{ik} S_{kj} \right)^2 \quad i = 1, \ldots, m, \quad j = 1, \ldots, n. \tag{2}$$

The General Mixture Problem and the special case of multivariate receptor modeling are ill posed problems. There are simply more unknowns than equations and thus there are many wildly different solutions that are all equally good in the least-squares sense. In a statistical way these problems are not identifiable. One approach to ill-posed prob-
5 lems is to impose conditions that add additional equations, which then define more realistic solutions to be closer to unique solution. The non-negativity conditions as additional conditions are imposed by the physical nature of the problem (Henry, 2001). Source composition and contributions must be non-negative but non-negativity conditions alone are not sufficient to give a unique solution. More constraints are needed
10 (Henry, 1987). Under certain, rather mild conditions, the data themselves can provide the needed constrains (Henry, 1997). This is how Unmix works.

Based on the multivariate factor analysis and the results in factor profiles and contributions, Paatero and Tapper (Paatero and Tapper, 1993; Paatero and Tapper, 1994; Paatero, 1997) established the advanced factor analysis method – positive matrix fac-
15 torization (PMF). Several features are incorporated in this model:

- weights data points by their analytical uncertainties,

- constrains factor loadings and factor scores to non-negative values and thereby minimizes the ambiguity caused by rotating the factors,

- uses weighted least-squares fits for data,

20 - expresses factor loadings in mass units, which allows factors to be used directly as source signatures,

- provides uncertainties for factor loadings and factor scores.

In PMF, the matrix $\mathbf{X}(n \times m)$ includes measured mass concentrations, and is represented as the sum of the product of $\mathbf{G}(n \times p)$ and $\mathbf{F}(p \times m)$ matrices and the resid-
25 ual matrix $\mathbf{E}(n \times m)$, where n is the number of samples, m is the number of chemical

species, and p is the number of independent source types. This model can give a solution that can be displayed in matrix form:

$$\mathbf{X} = \mathbf{G} \cdot \mathbf{F} + \mathbf{E}. \tag{3}$$

The object function Q that is to be minimized is defined as:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\varepsilon_{ij}/u_{ij} \right)^{2} \tag{4}$$

where u_{ij} is the uncertainty of the species j in a sample i and residuals ε_{ij} i.e. the portion of the measured concentration that cannot be explained by the model is defined by:

$$\varepsilon_{ij} = x_{ij} - \sum_{k=1}^{p} g_{ik} f_{kj}. \tag{5}$$

g_{ik} is the concentration contribution of source k to sample i and f_{kj} (source profiles) is the fractional concentration of species j in the emissions from source k.

In addition, non-negativity constraints should be fulfilled, meaning that all the elements in $\mathbf{G}$ and $\mathbf{F}$ are to be non-negative. The main process of the PMF is minimizing the Q value which is defined in the Eq. (4) as the sum of square of the residuals (ε_{ij}) weighted inversely with uncertainty (u_{ij}) of the data point (Polissar et al., 1998; Lee and Hopke, 2006). The procedure of Polissar et al. (1998) was used in this study to calculate uncertainties in the species concentrations. Briefly, for the data below detection limit (DL), the concentrations were replaced with the value DL/2 and the uncertainty was set as:

$$u_{ij} = \frac{5}{6} \times \text{DL}. \tag{6}$$

Printer-friendly Version

Interactive Discussion

For the missing data, concentrations were replaced by the geometric mean and the respective uncertainty was set at four times of this mean concentration. PMF provides uncertainties in factor profiles ($\sigma_{\mathrm{f}kj}$), defined as the standard deviation of 100 boot-strapping runs.

The solution of Eq. (4) is obtained by iteration until convergence is reached.

The output from the PMF model is a set of factors representing source profiles and estimates of their associated contributions to measured concentrations at the sampled receptor sites. Interpretation of the factors (i.e. allocation to names source types) has to be done by reference to information on source emissions, derived from literature and/or available measured data.

Numerous studies employing both the PMF and Unmix models have been done in recent years (Pekney et al., 2006; Poirot et al., 2001; Kim et al., 2004; Chen et al., 2007 in: Hegg et al., 2010). The models tend to be more or less in agreement with large data sets, but differences of a factor of two in the contribution of identified similar sources to particular samples are not uncommon (Hegg et al., 2010). Typically, Unmix resolves fewer factors than PMF and very importantly for our purposes – the nature of the factors resolved is much more dependent on the precise choice of input species than is the case for PMF (Maykut et al., 2003) Unmix will, in many instances, not yield a feasible solution if certain species are (or are not) included in the input and one commonly ends up running Unmix with fewer and different species than PMF even for the same data set (Hegg et al., 2010).

Paatero's positive matrix factorization (PMF) approach weights the data by the inverse of the measurement error for each observation. A major advantage of this approach is that the missing data can be included as observations with a large error. However, the minimization of **F** is still an ill-posed problem, or in other words, the model is not identifiable. Even the inclusion of the non-negative constraints does not provide an identifiable model. Paatero addresses this problem, which he named rotationally indeterminacy, by adding one or more user-selected parameters. Park et al. (2002) have used modern constrained minimization methods on F along with specific conditions,

Applying receptor models Unmix and PMF on real data set of elements in PM

D. Đorđević et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◄◄   ►►

◄   ►

Back   Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

e.g. each source composition must have at least one species absent from that source. Finally, Paatero has generalized F in a natural way to include the estimation of even more unknown parameters associated with spatial variations (Henry, 2002). Source apportionment by factors with explicit restrictions (SAFER) and Unmix finish the list of the selfmodeling curve resolution-based models. SAFER uses additional physical constrains on the source composition that must be supplied by the user. Linear programming is then applied to find the small regions that must contain the source points. Unmix uses the data to define edges, which are used to find the source points. Unmix can use additional physical constraints supplied by the user. At this time, this is limited to the identification of a species as a tracer, i.e. a species with effectively only one source. Currently, SAFER and Unmix (Henry, 2002, 2003) and Multilinear Engine ME-2 as a part of PMF 3.0 (Valentin et al., 2008)) are the frequently used methods that allow the additional physical constraints that can make use of the users' knowledge of the problem.

The main goal of this study is applying Unmix and PMF models on a data base that was previously submitted to other models; Principal Component Analysis (PCA) and Enrichment Factors (EA) have already indicated the main effects. Since Unmix and PMF are still under development, this analysis is useful for comparing results and evaluating models.

2   Materials and methods

Multivariate source apportionment models, Unmix and positive matrix factorization (PMF), often produce nearly the same source apportionment, however some investigations have shown that this is not always the case (Henry and Christensen, 2010). These models do not specify a minimum number of samples, but the stability of their solutions increases with the number of samples (Chen et al., 2007). In this study, we calculated sources composition and sources contributions of elements in PM using real data base. The real data set of 11 trace elements in particulate matter (PM) obtained in

Applying receptor models Unmix and PMF on real data set of elements in PM

D. Đorđević et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◄ | ►◄

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

252 observations was analyzed by Unmix 6.0 and PMF 3.0. The sampling site is situated only 10 m away from the coast of the Adriatic Sea. Samples of PM were subjected to gravimetric analysis for determination of total mass concentrations and subsequently to elemental analysis for Fe, Mn, Ti, Pb, Cr, Cu, Cd, Co, Ni, Hg and Se. Suspended particles were collected using a high-volume Aerosol Sampler, AQUERO model 400XT sampling system, on boron-silicate fibreglass filters every sixth day as a 24 h sample. The sampler was located in the town of Herceg-Novi (Fig. 8) 18°33″ N, 42°27″. The meteorological station is part of the MED POL program. The nearest road is located about 100 m north of the meteorological station. There are no significant grassy areas around the meteorological station, and there is no considerable construction work in progress. The terrain surrounding the receptor is rocky with some small areas of soil (Đorđević et al., 2004). Filters were digested with HNO_3. A Flame Atomic Absorption Spectrometry (F-AAS), Varian AAS-Spectr AA 55 instrument, was used to measure the concentrations of Cd, Co, Cr, Cu, Ni, Pb, Ti, Fe and Mn. The concentrations of Hg and Se were determined by the hydride vapor AAS method (HV-AAS) (Đorđević et al., 2005). The maximum expanded uncertainty of measurements for all elements was about 5 %.

3 Results of the Unmix model

The applied Unmix and PMF models were available on the EPA Internet site (www.epa.gov). Unmix and PMF used in this study do not limit the number of factors.

The following initial operations were subjected to the Unmix model data: *Suggest Exclusion*, *Initial Species*, *Additional Species* including SAFER and *Initial Points*. PM was chosen for the total and for the normalization.

The data was screened using the signal-to-noise ratio (Min S/N ratio) criteria higher than 2, estimated by Unmix. Only the component with S/N value greater than 2 will be used for sources estimation. The agreement between the true and estimated source contribution (Min R^2) was considered as well (Henry, 2003, EPA/600/R-07/089).

Figure 1 shows the comparison of measured and the predicted concentrations of PM and trace elements in PM through time series and Min R^2. Model Unmix did not calculate R^2 values for Cd, Co, Hg and Se and neither satisfactory solution included these elements. Min R^2 values are given in Table 1.

From statistical parameters displayed for each species, after input data and the following operations: *Suggest exclusion*, *Influential points*, *Initial species*, *Additional species* and *SAFER*, Unmix finds six combinations of species that give any kind of solution (Table 2). Min S/N for each principal component greater than 2 and Min R^2 greater than 0.8 of all combinations of elements estimated by Unmix were selected as good solutions that are in accordance with the Unmix criteria (Henry, 2003). Thirteen solutions in total were found, but four solutions satisfy the above criteria, signed in bold in Table 2. Parameters given for these solutions are also Significant/Strong Species in Sources (sigma-based).

Taking into account the calculated good solutions presented in Table 2, the Edges plots were done for these solutions (Fig. 2). In the first solution (combination of species Mn-Ti-Pb-Cr-Cu, 3 Sources Solution) the second and third source are well defined by many points on the y-axis while source 1 has just a few points on the x-axis.

The second solution (combination of species Fe-Mn-Ti-Pb, 3 Sources Solution) does not show good accumulation of points on the x and y-axes.

The third solution (combination of species Fe-Mn-Ti-Pb-Cr-Cu-Ni, 4 Sources Solution) shows the edges on the y-axis defined by many points for the third and the fourth source, but the x-axis has just a few points.

In the fourth solution (combination of species Fe-Mn-Ti-Pb-Cr-Cu, 3 Sources Solution) good accumulation of points are on the y-axis, for sources 2 and 3 while the x-axis has just a few points for source 1 (Fig. 2).

The source profile of the solutions chosen according to the criteria $S/N > 2$ and $R^2 > 2$ are given in Fig. 3.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

4 Results of the PMF model

At the first set up all elements are labeled as *Strong*, since (the signal/noise ratio) $S/N > 2$ for all of them. Based on input data statistics, residuals show bimodal distribution in the case of Ni, Mn, and Hg, so their uncertainties are increased labeling them as *Weak*. Selenium is excluded from the model because of a very small contribution and the correlation factor, while for cadmium more than 50 % of samples are below the detection limit.

The Q value represents the goodness-of-fit and assesses how the model fits the experimental data. Q_{true} is calculated taking into account all data points while Q_{robust} is calculated accounting for outlier points. Data with scaled residuals above 4 are regarded as outlier points.

The F peak function is used to rotate the data set, make fine tuning and improvement of the model in the case of data with high noise (positive values F peak) or clean data (negative values F peak). Normally, the default settings give satisfactory results, but in some cases subsequent adjustments are needed.

Bootstrapping is an advanced analysis that examines the stability of solutions of the tested models. The bootstrap method is essentially based on resampling methods in which "new" data sets that are consistent with the original data are generated. Each "new" data set (which is essentially a subset of the original database), is decomposed into profile and contribution matrices, and the resulting profile and contribution matrices are compared with the base run (Eberly, 2005), giving the distribution for each species to evaluate the stability of the solution.

To ensure the robustness of statistics, 300 bootstrap runs were performed, while the default value of the minimum correlation (R value) of 0.60 was used.

The application of positive matrix factorization to solve the number and profile of the sources applied to a database of measured concentrations of elements (Đorđević et al., 2005) resulted in obtaining possible solutions for 3, 4 and 5 sources. For 6 or more sources the model does not find the convergence of the functions Q, which implies that

the model did not find any minima. Varying simulation conditions did not contribute to significant improvement, even when the uncertainty is significantly increased. Therefore possible solution should be sought among three possible cases. The lack of a solution with a larger number of sources was reflected in the inability to use the IM – the maximum individual column mean and IS – the maximum individual column standard deviation (Lee et al., 1999), which would be helpful to find the true number of sources.

Table 3 shows the categories of elements and the R^2 values for each of the three possible solutions.

Each of the possible solutions obtained by PMF analysis will be considered. Figure 4 shows F peak strengths for 3 sources solution (Fig. 4a), 4 sources solution (Fig. 4b) and 5 sources solution (Fig. 4c).

3 Sources Solution: the relatively good correlation was obtained only for Cr and Pb, while bimodal distribution is still present in the case of Co, Ti and Fe. Also, significant outliers are present in the model. In addition, G-Space plots show considerable interaction between the sources 1 and 3.

Rotational ambiguity, which was found between the sources 1 and 3, decreases when the value of Strength factor reaches -1.2 (Fig. 4a). This is mostly reflected in the increase of Ti concentration in the source 2.

However, a small degree of correlation between the model and database indicates that the model with three sources is insufficient to adequately describe a number of sources.

In this case, only Co, Ni and Fe show relatively good interquartile range of about 20 %, while other species show considerable variation and therefore represent a less stable solution. This is especially pronounced in the case of Hg, Cr and Mn. Also, in some cases (Hg, Ti) base run values are not within the interquartile range in the bootstrapping of results. This is probably a consequence of assuming the model with only three sources. Profiles of sources are given in Fig. 5.

4 Sources Solution: the model with four sources shows a significantly better correlation with measured concentrations of elements. Although the agreement of time series for Ti and Cr is excellent ($R^2 > 0.95$), and for Pb satisfactory ($R^2 = 0.70$), in the case of other elements there are still episodes with very high concentrations that this model cannot fit. It should be noted that Cu shows very good agreement between the predicted and observed concentrations, but the existence of outliers have reduced the correlation to 0.34. A small degree of correlation in the case of Co is the result of a significant number of measurements below the detection limit.

Bimodal distribution is still present in the case of Ni and Hg. G-Space plot shows some interaction between sources 1–3, 2–3, 3–4.

For a model with four sources, rotational ambiguity disappears when the *F* peak strength reaches −0.8 (Fig. 4b). This rotation is mostly reflected in the increase of Ti content in source 1, and largely in sources 2 and 3. On the other hand, this may just mean that the content of titanium in this solution is divided among several sources. It is necessary to consider these results carefully and determine whether there is justification for it to be included in further consideration of the composition of the sources.

The existence of significant unmapped results indicates that only the relative stability of the obtained solutions is achieved. Interquartile ranges in the case of Fe, Pb, Cu and Cr are about 20 %, while in the case of other species this range is much higher indicating the instability of the solution. Base run values which are not within the interquartile range in the bootstrapping of the results are, in the case of Cu, Mn, Pb and especially Hg, calculated by the model only in the fourth source.

5 Sources Solution: The model and data from the database show agreement (R^2) over 90 % for Cr, Ti, Fe and Pb, while just over 50 % for Mn. The model also fits Cu real data very well, and the correlation of 0.47 is caused by significant outliers that are related to individual episodes of Cu emissions. In spite of the increased uncertainty Mn, Ni, Co and Hg show a lack of fit.

G-Space plot only shows some rotational ambiguity in the case of sources: 2–5, 3–5 and 4–5.

The F peak in the range −2 to 2 (Fig. 4c) showed the most impact on the sources of Cu, especially at higher strength values, while the ambiguity between the sources mentioned above still exists. The peak F curve is generally symmetrical in the examined interval.

In the case of five sources there are also unmapped results, suggesting a reduced stability of the solution. In general, the most stable solutions are obtained for those elements that are present in the source with the highest percentage (interquartile range then is approximately 15 %). This is the case for Fe, Pb, Ti. A slightly worse result of the bootstrap analysis is obtained for Cu, Mn, Cr, Co and Ni (interquartile ranges are equal or higher than 20 %). The least stable solution is for Hg with considerable dispersion in the bootstrap analysis solutions. However, it should be noted that the interquartile and R^2 value from Table 3 for the corresponding elements are in agreement.

5 Discussion

In our previous work (Đorđević et al., 2005) we applied the PCA method on this data set and 4 significant groups of sources contributions were found. The dominant factor (F1) is represented by the contribution sources of Fe, Mn, Ti, Cu and Pb. The following contribution sources were identified: re-suspension combined with the transport of Saharan dust from North Africa (Fe, Mn, Ti), re-suspended Saharan dust that had previously settled (Fe, Mn, Ti) and settled combustion products, originating from traffic, and probably some local stationary source (Cu, Pb). The remaining three factors represent the following combinations F2 by Cr and Ni, F3 by Cd and Se and F4 by Hg and Co. Also, the Enrichment Factors model revealed that in the region of the investigated receptor, the main contribution source of Fe, Mn and Ti is the process of local re-suspension and that local re-suspension has no influence on the content of Se in the atmospheric aerosol. The re-suspension is the dominant emission source of Cd from the south-southeast direction from the nearby peninsula (Luštica) but this source is not permanent. The re-suspension also contributes to the contents of Pb, Ni, Cr, Cu

and Co as a consequence of their deposition in the vicinity of the receptor after being emitted from anthropogenic sources in the region (Đorđević et al., 2005).

The Unmix model excluded the following elements from the calculation: Cd, Co, Hg and Se. In the first satisfactory solution for Mn-Ti-Pb-Cr-Cu combination Unmix found 3 sources (Table 2, Fig. 3a). Pb is strong in the first source and this source can be attributed to traffic. In the second source Cr and Cu are strong and Ti and Mn are significant. This source can be re-suspension of elements previously settled from anthropogenic sources. In the third source neither element is strong or significant.

In the second satisfactory solution for Fe-Mn-Ti-Pb combination of elements it also found 3 sources (Table 2, Fig. 3b). This solution has the best values of Min R^2 and Min S/N compared to all combinations. The first and the third source contain Pb which is a tracer for traffic. In the third source Pb is strong, and it is reasonable to associate this source with traffic, while the first source could be local re-suspension. The second source in this combination could be a long range transport of Saharan dust since it contains crustal elements.

In the third and the fourth solution, the sources where Pb is strong can be attributed to traffic; namely, source 3 for Fe-Mn-Ti-Pb-Cr-Cu-Ni combination (Table 2, Fig. 3c) and source 1 for Fe-Mn-Ti-Pb-Cr-Cu combination (Table 2, Fig. 3d). Another source in which Pb is present as significant but not strong could be re-suspension. Source 4 for Fe-Mn-Ti-Pb-Cr-Cu-Ni combination and source 2 for Fe-Mn-Ti-Pb-Cr-Cu combination could be attributed to re-suspennsion, probably from various locations depending on wind directions. The contribution sources of Fe, Mn, Ti, Cu and Pb are re-suspension of previously settled Saharan dust (Fe, Mn and Ti) and combustion products from traffic, combined with a long-range transport of Saharan dust from North Africa. Factors containing Cr and Ni indicate the existence of an anthropogenic emission source in the region. More precisely, in the northern segment at a distance of 100 km from the receptor, there is a slag heap, as well as a Coal Fired Power Plant – in Pljevlja and the Ironworks Nikšić (Đorđević et al., 2005).

When discussing the number and origin of pollution sources, it is preferred to take into account the real situation on the field. In this case the following sources that contribute to the overall PM deposition can clearly be predicted: marine aerosols, traffic, re-suspension from the ground, probably some local stationary source, as a shipyard located in the vicinity. Based on these obvious sources, PMF analysis solution with only three sources is exempt from further consideration.

In the case of PMF solution of five sources it may be noted that source No. 5 (Fig. 7), in which Co, Cu, Ni and Mn are present, can be described rather as a splitting factor than as a separate source. The most realistic solution that is imposed upon a detailed analysis is the solution with four sources (Fig. 6).

Identification of sources was carried out and it agrees with the results of the Enrichment Factors analysis well (Đorđević et al., 2005). The F peak profiles shown in Fig. 6 in rotation of data set for −0.8, increase the contents of Ti, in the case of sources 2 and 4.

Source 1 has been identified as re-suspension in combination with the long-range transport of Saharan dust. Figure 8 shows that the prevailing wind directions are over open sea.

Source 2 is attributed to the re-suspension coming from the direction of the nearby peninsula (Luštica), indicated in our previous work. Titanium found by F peak is in better accordance with the Enrichment Factors analysis (Đorđević et al., 2005).

Source 3 corresponds to the composition of the particles that come from some anthropogenic source.

Source 4 with the highest content of Pb, is characteristic for urban traffic. F peak is increasing the value for Ti which is in better agreement with the traffic profile.

6   Conclusions

In this study we have used concentrations of elements in PM measured in the urban area of a coastal city in the south east of the Adriatic Sea. The real data set of eleven

trace elements in PM obtained in 252 observations was analyzed by Unmix 6.0 and PMF 3.0. By applying Unmix to the real data set we found four satisfactory solutions in terms of modeling. The PMF model has given 3 possible solutions and by further analysis the best solution of four sources was selected. F peak refinement was enabled to find a more realistic solution. This solution indicated re-suspension and traffic as dominant source of contribution. All of these solutions are in accordance with the Enrichment Factors shown in our previous work.

References

Chen, L.-W. A., Watson, J. G., Chow J. C., and Magliano, K. L.: Quantifying PM$_{2.5}$ Source Contributions for the San Joaquin Valley with Multivariate Receptor Models, Environ. Sci. Technol., 41, 2818–2826, 2007.

Đorđević, D., Vukmirović, Z., Tosić, I., and Unkasević, M.: Contribution of dust transport and resuspension to particulate matter levels in the Mediterranean atmosphere, Atmos. Environ., 38, 3637–3645, 2004.

Đorđević, D., Mihajlidi-Zelić, A., and Relić, D.: Differentiation of the contribution of local resuspension from that of regional and remote sources on trace elements content in the atmospheric aerosol in the Mediterranean area, Atmos. Environ., 39, 6271–6281, 2005.

Eberly, S.: EPA PMF 1.1 user's guide, http://www.epa.gov/heasd/products/pmf/users_guideold.pdf, Prepared by the US Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, 2005.

Hegg, D. A., Covert, D. S., Jonsson, H. H., and Woods, R. K.: The contribution of anthropogenic aerosols to aerosol light-scattering and CCN activity in the California coastal zone, Atmos. Chem. Phys., 10, 7341–7351, doi:10.5194/acp-10-7341-2010, 2010.

Henry, R. C.: Current Factor Analysis Receptor Models are Ill-Posed, Atmos. Environ., 21, 1815–1820, 1987.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Henry, R. C.: History and Fundamentals of Multivariate Air Quality Receptor Models, Chemometr. Intell. Lab., 37, 525–530, 1997.

Henry, R. C.: UNMIX Version 2.4 Manual, US Environmental Protection Agency, Research Triangle Park, NC, 2001.

Henry, R. C.: Multivariate receptor models-current practice and future trends, Chemometr. Intell. Lab., 60, 43–48, 2002.

Henry, R. C.: Multivariate receptor modeling by N-dimensional edge detection, Chemometr. Intell. Lab., 65, 179–189, 2003.

Henry, R. C. and Christensen, E. R.: Selecting an Appropriate Multivariate Source Apportionment Model Result, Environ. Sci. Technol., 44, 2474–2481, 2010.

Kim, E., Hopke, P. K., Larson, T. V., and Covert, D. S.: Analysis of ambient particle size distributions using Unmix and positive matrix factorization, Environ. Sci. Technol., 38, 202–209, 2004.

Lee, E., Chan, C. K., and Paatero, P.: Application of positive matrix factorization in source apportionment of particulate pollutants in Hong Kong, Atmos. Environ., 33, 3201–3212, 1999.

Lee, J. H. and Hopke, P. K.: Apportioning sources of $PM_{2.5}$ in St. Luis, MO using speciation trends network data, Atmos. Environ., 40, S360–S377, 2006.

Lee, J. H., Lim, J. M., Kim, K. H., Chung, Y. S., and Lee, K. Y.: Trace element levels of aerosols at an urban area of Korea by instrumental neutron activation analysis, J. Radioanal. Nucl. Ch., 256, 553–560, 2003.

Lee, D., Balachandran, S., Pachon, J., Shankaran, R., Lee, S., Mulholland, J. A., and Russell, A. G.: Ensemble-Trained $PM_{2.5}$ Source Apportionment Approach for Health Studies, Environ. Sci. Technol., 43, 7023-7031, 2009.

Maykut, N. N., Lewtas, J., Kim, E., and Larson, T. V.: Source apportionment of $PM_{2.5}$ at an urban IMPROVE site in Seattle, Washington, Environ. Sci. Technol., 37, 5135–5142, 2003.

Norris, G. A., Vedantham, R., and Duvall, R. M.: EPA UNMIX 6.0 User Guuide, EPA/600/R-07/089 (NTIS PB2007-112630), US Environmental Protection Agency, Washington, DC, 2007.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, Chemometr. Intell. Lab., 37, 23–35, 1997.

Paatero, P. and Tapper, U.: Analysis of different modes of factor analysis as least square fit problem, Chemometr. Intell. Lab., 18, 183–194, 1993.

Applying receptor models Unmix and PMF on real data set of elements in PM

D. Đorđević et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

| ◄◄ | ►► |
| ◄ | ► |

Back | Close

Full Screen / Esc

Paatero, P. and Tapper, U.: Positive Matrix Factorization: a non-negative factor model with optimal utilization of error estimates of data values, Environmetrics, 5, 111–126, 1994.

Park, E. S., Spiegelman, C. H., and Henry, R. C.: Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models, Environmetics, 13, 775–798, 2002.

Pekney, N. J., Davidson, C. I., Robinson, A., Zhou, L., Hopke, P., Eatough, D., and Rogge, W. F.: Major source categories for PM$_{2.5}$ in Pittsburgh using PMF and UNMIX, Aerosol Sci. Tech., 40, 910–924, 2006.

Poirot, R. L., Wishinski, P. R., Hopke, P. K., and Polissar, A. V.: Comparative application of multiple receptor methods to identify aerosol sources in Northern Vermont, Environ. Sci. Technol., 35, 4622–4636, 2001.

Polissar, A. V., Hopke, P. K., Paatero, P., Malm, W. C., and Sisler, J. F.: Atmospheric aerosol over Alaska 2. Elemental composition and sources, J. Geophys. Res., 103, 19045–19057, 1998.

US Environmental Protection Agency, EPA Unmix 6.0 Model, http://www.epa.gov/heasd/products/Unmix/Unmix.htm, last access: June 2007.

US Environmental Protection Agency, EPA Positive Matrix Factorisation 3.0, http://www.epa.gov/heasd/products/pmf/pmf_registration.htm.htm, last access: July 2008.

Valentin, A. L., Rami Alfarra, M., Urs, B., Brigitte, B., Christoph, H., Sönke, S., Miriam, N. W., Lukas, W., Silke, W., Alexandre, C., Puxbaum, H., and Andre, S. H. P.: Source Attribution of Submicron Organic Aerosols during Wintertime Inversions by Advanced Factor Analysis of Aerosol Mass Spectra, Environ. Sci. Technol., 42, 214–220, 2008.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

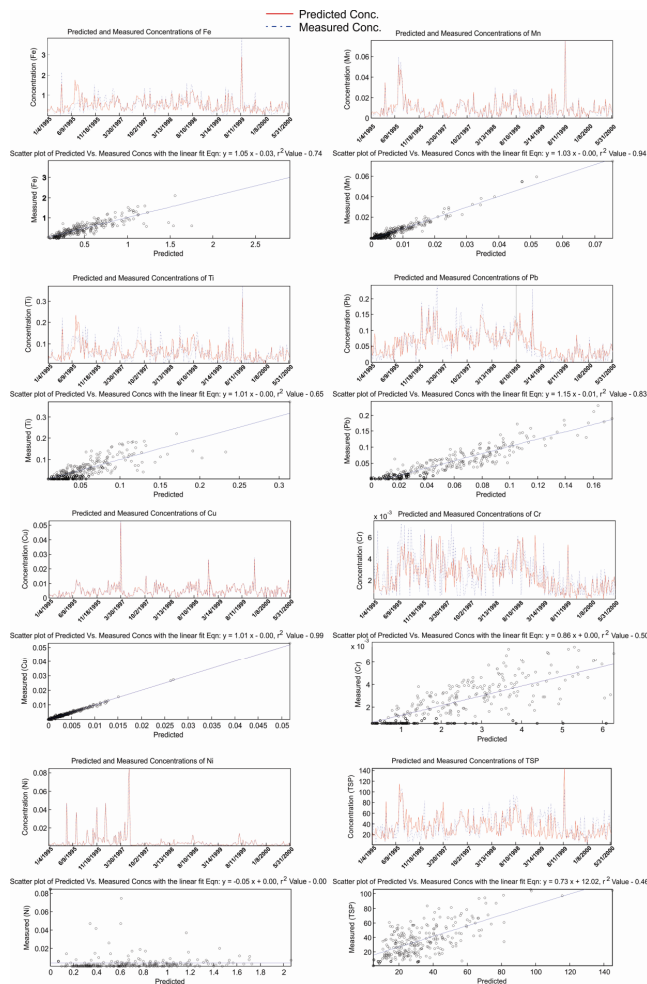Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. R^2 values obtained by Unmix of measured and the predicted concentrations of PM and trace elements in PM.

	PM	Fe	Mn	Ti	Pb	Cr	Cu	Ni
R^2	0.46	0.74	0.94	0.65	0.83	0.50	0.99	0.00

Table 2. All combinationas of elements for solutions obtained by calculation by Unmix.

Combination of species	Number of sources	Min R^2	Min S/N	Significant/strong species in sources (σ-based)
Mn-Ti-Pb-Cr-Cu	**3**	**0.84**	**2.49**	Source 1: Strong[*] – Pb;
	4	0.89	1.94	Source 2: Strong[*] – Cr, Cu, Significant[**] – Ti, Mn;
	5	0.90	1.59	Source 3: Strong[*] – None, Significant[**] – None
Cr-Cu-Pb-Ti-Mn-Se-Cd-Co-Fe	3	0.68	2.41	
	4	0.76	2.13	
Cu-Ti-Fe-Mn-Pb-Cr-Hg-Se	2	0.56	2.29	
Fe-Mn-Ti-Pb	**3**	**0.90**	**2.85**	Source 1: Strong[*] – None, Significant[**] – PM, Pb, Fe
				Source 2: Strong[*] – None Significant[**] – PM, Fe, Mn;
				Source 3: Strong[*] – Pb, Mn, Significant[**] – Ti, Fe
Fe-Mn-Ti-Pb-Cr-Cu-Ni	3	0.76	2.20	Source 1: Strong[*] – Cr, Ni, Significant[**] – None
	4	**0.83**	**2.18**	Source 2: Strong[*] – None, Significant[**] – Cu;
	5	0.89	1.67	Source 3: Strong[*] – Pb, Significant[**] – PM, Cr, Cu, Ti, Fe, Mn;
				Source 4: Strong[*] – Ti, Significant[**] – PM, Cr, Pb, Fe
Fe-Mn-Ti-Pb-Cr-Cu	**3**	**0.83**	**2.57**	Source 1: Strong[*] – Pb, Significant[**] – Cu;
	4	0.88	1.97	Source 2: Strong[*] – None, Significant[**] – PM, Cr, Pb, Ti, Fe;
	5	0.90	1.62	Source 3: Strong[*] – None, Significant[**] – PM, Cr, Cu, Ti, Fe, Mn

[*] Source composition ≥ 1 σ; [**] source composition ≥ 2 σ. The satisfactory solutions are in bold.

Table 3. R^2 values obtained by PMF of measured and the predicted concentrations.

Species	Category	R^2		
		3 sources	4 sources	5 sources
Cr	Strong	0.617	0.980	0.998
Ti	Strong	0.381	0.962	0.959
Fe	Strong	0.401	0.472	0.942
Pb	Strong	0.695	0.701	0.905
Mn	Weak	0.394	0.340	0.528
Cu	Strong	0.345	0.337	0.473
Ni	Weak	0.027	0.031	0.025
Co	Weak	0.006	0.013	0.018
Hg	Weak	0.002	0.005	0.003

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Fig. 1. Predicted and measured concentrations.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
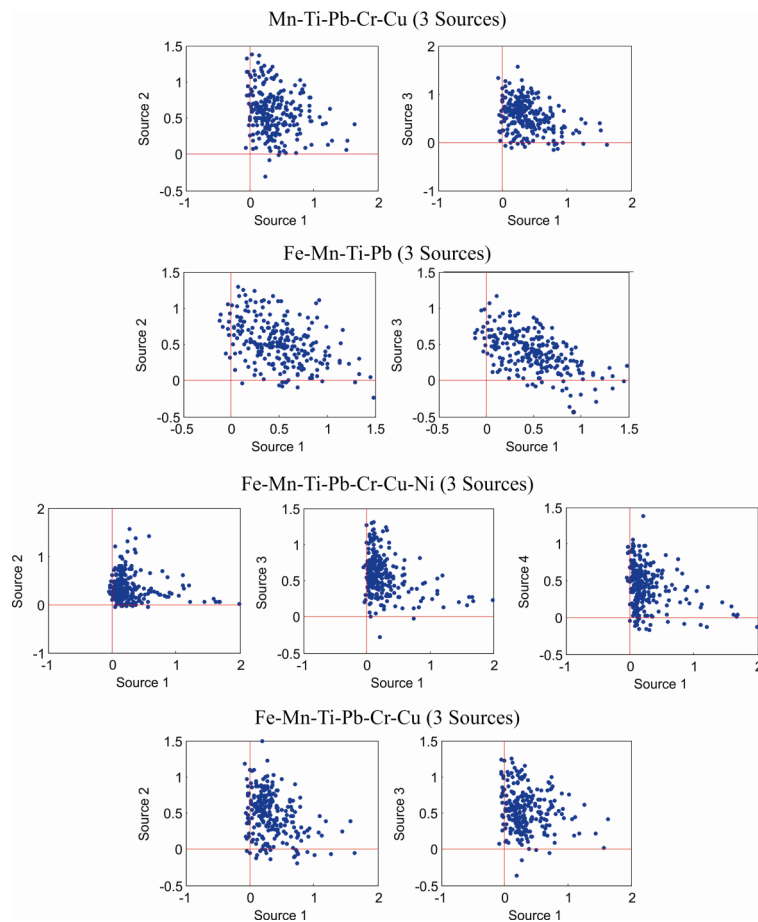
Fig. 2. Edge plots for chosen solutions that satisfy the conditions of Min S/N and Min R^2.
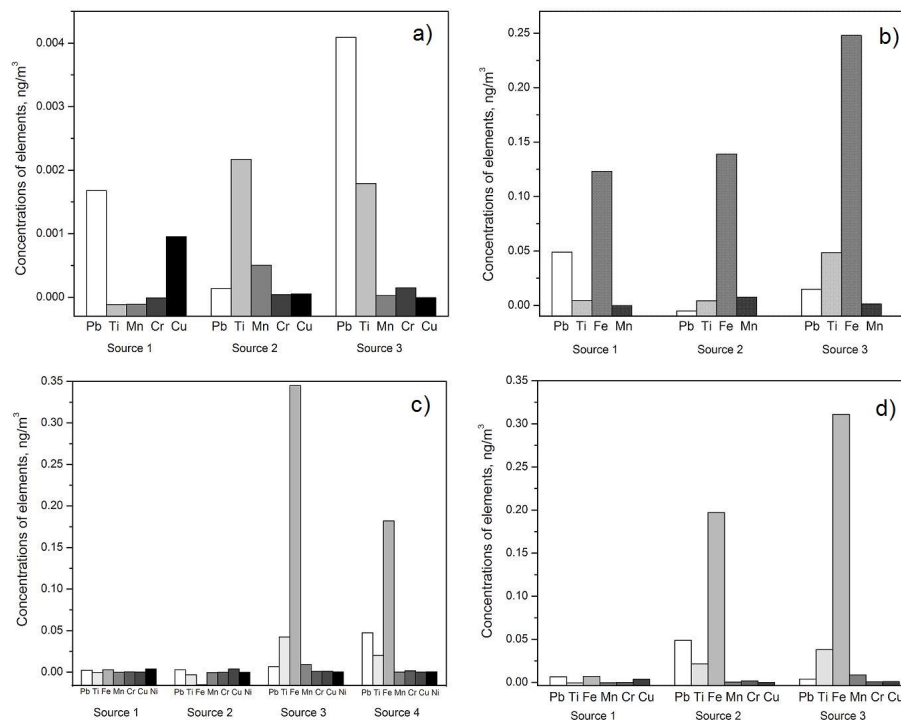
Fig. 3. Source profiles for selected solutions that are in accordance with the Unmix criteria.

D. Đorđević et al.
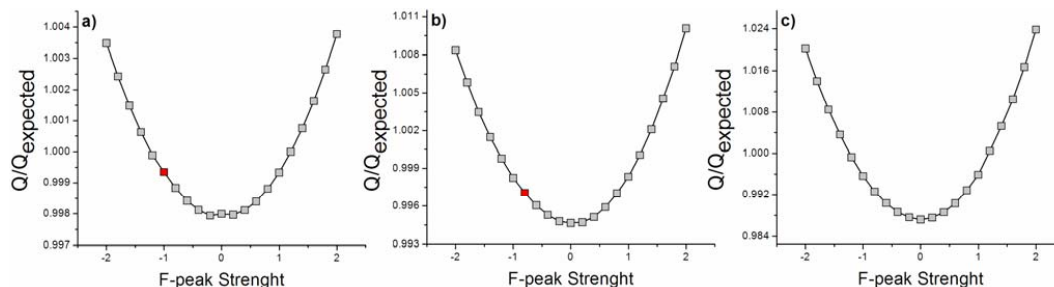
Fig. 4. *F* peak analysis for three **(a)**, four **(b)** and five **(c)** source solutions. The red mark represents the value of *F* peak Strength, at which the rotational ambiguity disappears.
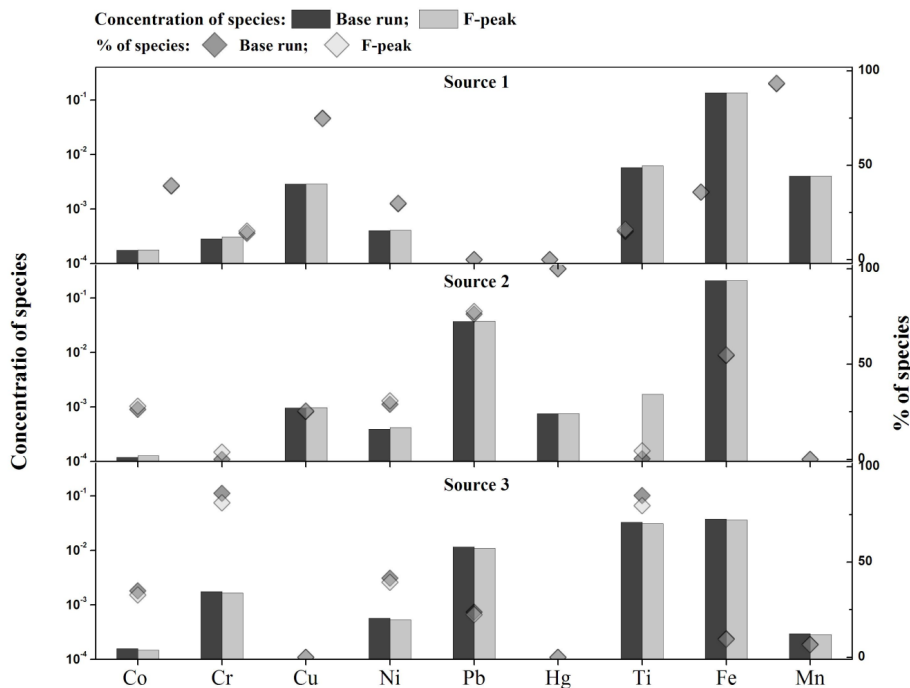
Fig. 5. Profiles in the case of three sources solutions. Comparison of base run profile and *F* peak run profile with the strength of −1.2 (disappearance of rotational ambiguity).
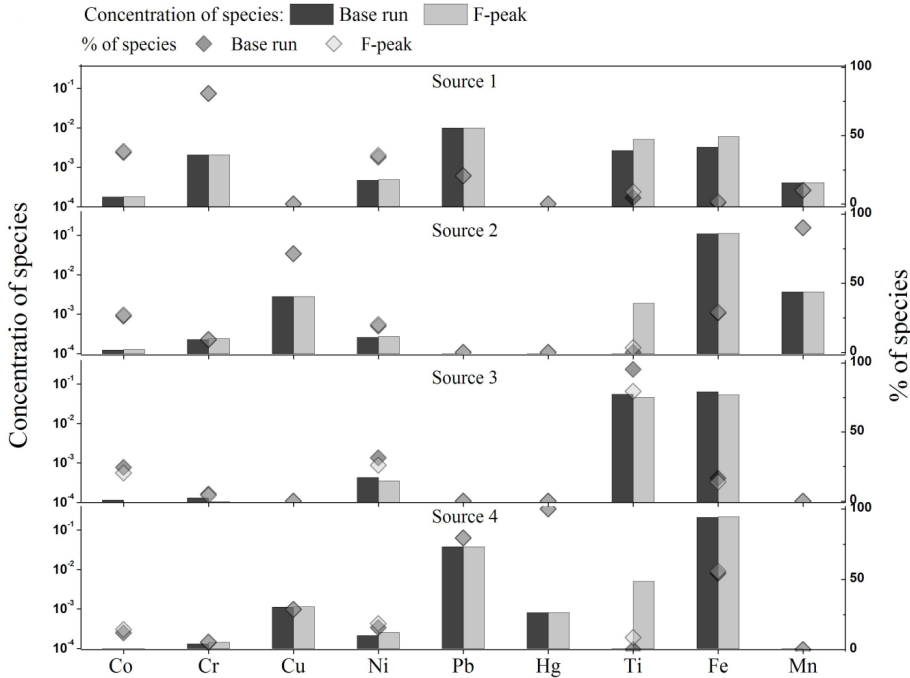
Fig. 6. Profiles in the case of four sources solutions. Comparison of base run profile and *F* peak run profile with the strength of −0.8 (disappearance of rotational ambiguity).
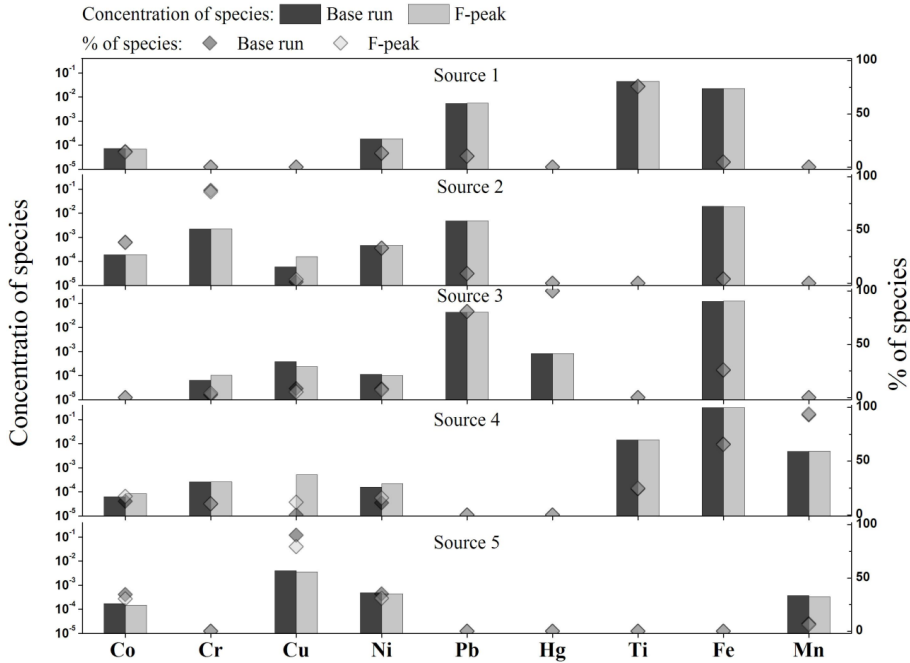
Fig. 7. Profiles in the case of five sources solutions. Comparison of base run profile and F peak run profile with the strength of −2.0 where it can be seen that F peak Strength does not affect the existing rotational ambiguity.
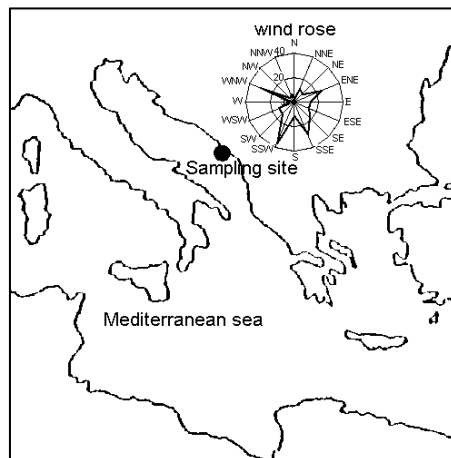
Fig. 8. Sampling site and prevailing wind directions.